HHS Public Access

Author manuscript

Nat Ecol Evol. Author manuscript; available in PMC 2021 August 14.

Published in final edited form as:

Nat Ecol Evol. 2020 September ; 4(9): 1239–1246. doi:10.1038/s41559-020-1232-4.

Birth of a pathway for sulfur metabolism in early amniote evolution

Marco Malatesta[1], **Giulia Mori**[1], **Domenico Acquotti**[2], **Barbara Campanini**[3], **Alessio Peracchi**[1], **Parker B. Antin**[4], **Riccardo Percudani**[1,*]

[1]Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Italy.

[2]Centro Interdipartimentale Misure “Giuseppe Casnati”, University of Parma, Italy.

[3]Department of Food and Drug, University of Parma, Italy.

[4]Department of Cellular and Molecular Medicine, University of Arizona, Tucson, USA

Abstract

Among amniotes, reptiles and mammals are differently adapted to terrestrial life. It is well appreciated that terrestrialization required adaptive changes of vertebrate metabolism, particularly in the mode of nitrogen excretion. However, the current paradigm is that metabolic adaptation to life on land did not involve synthesis of enzymatic pathways *de novo* but repurposing of existing ones. Here, by comparing the inventory of pyridoxal phosphate-dependent enzymes (PLPome) in different amniotes, we identify *in silico* a pathway for sulfur metabolism present in chick embryos but not in mammals. Cysteine lyase (CL) contains heme and PLP cofactors and converts cysteine and sulfite into cysteic acid and hydrogen sulfide. A specific cysteic acid decarboxylase (CAD) produces taurine while hydrogen sulfide is recycled into cysteine by cystathionine beta-synthase (CBS). This reaction sequence enables the formation of sulfonated amino acids during embryo development in the egg at no cost of reduced sulfur. The pathway originated around 300 million years ago in a proto-reptile by CBS duplication, CL neofunctionalization, and CAD co-option. Our findings indicate that adaptation to terrestrial life involved innovations in metabolic pathways and reveal the molecular mechanisms by which such innovations arose in amniote evolution.

Amniotes split ~320 million years ago soon after their origin into two lineages that dominate animal life on land: Sauropsida (Reptilia) including turtles, lizards, snakes, crocodiles, and birds, and Synapsida whose extant representatives are mammals[1–3]. Before separation, amniotes had developed few derived characters, most notably a protective and nourishing

*Correspondence to: riccardo.percudani@unipr.it.

Author contributions

R.P. and M.M. performed bioinformatics, M.M. performed protein production, M.M. and G.M. performed protein characterization, P.B.A. supervised *in situ* hybridization, D.A. supervised NMR spectroscopy, B.C. supervised fluorescence spectroscopy, A.P. conceived and curated B6db, R.P. conceived the study and wrote the manuscript with contributions by P.B.A., G.M., M.M., B.C., and A.P.

Competing interests

The authors declare no competing interests

structure for embryo development[4]. After separation, reptiles and mammals evolved distinct adaptations to terrestrial life[5]. These adaptations are less well understood at the molecular level. It has long been known that the two amniote clades evolved patterns of nitrogen elimination (uricotelism and ureotelism) suitable for their survival and development in terrestrial ecosystems[6,7]. Such modifications in nitrogen metabolism involved the use of old pathways for a new purpose[8,9], supporting the notion that terrestrialization did not require novel enzymes and metabolic pathways. Consistent with this view, no evidence has been found for enzymatic pathways that originated in early amniote evolution[10–12]. Whole genome comparisons, however, have revealed a number of genes that are not shared between the two classes of amniotes -more than two thousand in the case of *Gallus gallus* versus *Homo sapiens*[13]. Whether these genetic differences correspond to innovations in molecular processes and pathways is largely unknown. Conversely, there is evidence for metabolic pathways present in sauropsids but absent in mammals for which the genes have not been identified.

According to ^{35}S radiotracer experiments[14,15], chicken embryos synthesize sulfonated amino acids through incorporation of sulfite into cysteine with release of H_2S. During development, this activity (cysteine lyase; EC 4.4.1.10) is first observed in germ layer cells and increases with the differentiation of the yolk-sac endoderm[16]. The presence of cysteine lyase was confirmed in other sauropsids, but not in mammals[17]. The partially purified enzyme[18] was found to be dependent on pyridoxal 5'-phosphate (PLP). This activity, however, has never been reported, even as a side reaction, for any known PLP-dependent protein. In embryonated eggs, sulfite or sulfate ^{35}S is finally incorporated into taurine[14,15], a sulfonated amino acid (2-aminoethane sulfonic acid) with a vital role in vertebrates during development and adult life[19,20]. In mammals, taurine is synthesized in a pathway starting with cysteine/cysteamine oxidation[19,21,22]. However, during fetal development in humans[23], cats[20], and mice[24], taurine is provided by maternal transfer. In view of previous evidence, a pathway for taurine biosynthesis involving cysteine lyase and oxidized sulfur should exist in embryonated chicken eggs. The understanding of this pathway, however, has been limited by the lack of knowledge of its molecular components. Such missing pieces of information in metabolic reactions, or "pathway holes"[25], represent a substantial portion (10–30% depending on the organism) of pathway databases (e.g. Kegg: https://www.genome.jp/kegg; Metacyc: https://metacyc.org). With the availability of complete genomes, the genes responsible for such unassigned functions can be searched in a finite list of sequences by bioinformatics. There are several bioinformatics methods for the prediction of gene-trait associations[25,26]. Here we devised a procedure to limit the search space based on information of the reaction mechanism and dependency on a particular cofactor.

Results

To identify this pathway, we assumed that *Gallus gallus* has a gene encoding a PLP-dependent enzyme that is absent in mammals. Enzymes that depend on PLP (the active form of vitamin B6) are remarkable for their evolvability: they catalyze a wide variety of reactions, but have a limited number of evolutionary origins, making it possible to identify an organism's PLPome *in silico*. We initially compared the *Gallus gallus* and *Homo sapiens* PLPomes (Fig. 1a) with the genome analysis tool of B6db[27]. This side-by-side

comparison (Extended Data Fig. 1) showed that these species encode only ~50 of the >300 known families contained in the database. Most of the proteins in the two amniotes have a 1:1 orthologous relationship. However, 8 human and 4 chicken proteins do not have a correspondence in the other species (Fig. 1a). One of the 4 proteins uniquely present in *Gallus* (XP_015151050; predicted threonine aldolase) could be excluded as comparisons that included other species revealed that this gene is generally present in mammals. Detailed examination of the remaining proteins identified a single strong candidate for the sought function. In the B6db classification, the XP_015156382 protein (gene: LOC418544) was assigned to the same family as cystathionine beta-synthase (CBS), an enzyme catalyzing β-replacement reactions similar to cysteine lyase. In particular, the main cysteine lyase activity (reaction 1, Fig. 1b) resembles the serine hydro-lyase activity of CBS (reaction 4, Fig. 1c), while a secondary activity of cysteine lyase, i.e. formation of lanthionine (reaction 2, Fig. 1b), resembles the main CBS activity (reaction 3, Fig. 1c). We used 3D structures of experimentally validated CBS proteins and structural models of homologous *Gallus* sequences to analyze the conservation of residues lining the active site. The protein annotated as CBS (XP_015156364) had perfect conservation in this region, while the protein annotated as CBS-like (XP_015156382) showed conservation of the catalytic lysine (K119), but several non conservative substitutions of active site residues (Fig. 1d). Chicken embryo gene expression data in the GEISHA database[28] suggested LOC18544 expression at an extra-embryonic location consistent with the reported cysteine lyase activity[16]. In view of the evidence of the bioinformatics analysis we undertook the characterization of the protein encoded by LOC418544. Based on the observations described below, we named this gene cysteine lyase (CL).

The *Gg*CL protein sequence has the same level of similarity with either *Gg*CBS or *Hs*CBS (65.1% and 64.6% identity), less than the similarity between the two CBS sequences (75.3% identity) (Extended Data Fig. 2a). Domain analysis predicts for *Gg*CL an architecture similar to CBS[29], with a N-terminal heme-binding domain, a central PLP-binding domain, and two C-terminal CBS repeats (Fig. 1e). As previously reported for CBS[30], soluble expression could be achieved with a truncated protein lacking the regulatory CBS repeats (aa 1–396, Fig. 1e). Recombinant *Gg*CL was produced in *E. coli* and purified to homogeneity as an orange protein (Extended Data Fig. 2b,c). Conservation of residues for heme and PLP coordination (Extended Data Fig. 2a,d) suggests that CL has maintained the ability to bind these cofactors. The absorbance spectra of *Gg*CL showed the typical Soret peak of heme-binding proteins (Extended Data Fig. 2e). The presence of PLP was not apparent in the absorbance spectrum due to the dominant signal of heme. However, the fluorescence emission spectrum upon excitation at 412 nm showed a peak centered at 510 nm (Extended Data Fig. 2f) attributable to the ketoenamine tautomer of bound PLP[31].

We monitored spectrophotometrically the *Gg*CL activity by trapping *in situ* generated H_2S with lead acetate to form lead sulfide (PbS), a dark compound. H_2S release was observed with cysteine alone, and was much faster in the presence of sulfite (Extended Data Fig. 2g), suggesting that *Gg*CL catalyzes both cysteine lyase reactions (see Fig. 1b). In the absence of lead acetate, these reactions, albeit conducted with just micromoles of reagents, could be perceived for their rotten egg odor. H_2S is a volatile molecule with a role as gaseous messenger in vertebrate[32]. The dependence of reaction velocity on substrate concentrations

followed Michaelis-Menten kinetics (Extended Data Fig. 2h–i) with Fitting to the MM equation gave a k_{cat} value of 17.32 ± 1.05 s^{-1} and K_M values of 12.75 ± 2.16 mM for cysteine and 0.096 ± 0.018 mM for sulfite. Formation of the cysteine reaction product was directly followed by time-resolved ^{1}H NMR spectrometry. In the presence of excess sulfite, cysteine was completely converted into cysteic acid (CA) (Fig. 1f), whereas a partial conversion into lanthionine was observed in the absence of sulfite (Extended Data Fig. 2j). *Gg*CL showed no activity with serine and homocysteine, or serine and H_2S (Extended Data Fig. 3a,b), indicating that it is unable to catalyze the CBS reactions (see Fig. 1c and Extended Data Fig 3c). While incompetent in the β-replacement of serine, *Gg*CL is able to abstract the serine alpha proton (Extended Data Fig. 3d–e), i.e. to complete the first step of the reaction mechanism. By contrast, recombinant *Gg*CBS was found to be able to catalyze the CBS reactions, including the β-replacement of serine with H_2S to form cysteine (Fig. 1g).

The CA product of the CL reaction could be converted into taurine in a single step. This reaction, which involves α-decarboxylation of an amino acid, is presumably catalyzed by a PLP-dependent enzyme. Our PLPome comparison did not reveal a putative decarboxylase uniquely present in *Gallus*. However, *Gallus* has a *bona fide* ortholog (XP_025001259; see Fig. 1a and Extended Data Fig. 1) of mammalian cysteine sulfinic acid decarboxylase (CSAD), a protein reportedly able to catalyze CA decarboxylation, albeit with lower efficiency with respect to its CSA substrate[33]. The XP_025001259 protein, containing the typical domain of PLP-dependent decarboxylases (Fig. 1h), was produced in *E. coli* as a PLP-bound protein, with a prevalent enolimine tautomer of the cofactor (Extended Data Fig. 4a–b). Surprisingly, not only the *Gallus* protein was able to efficiently catalyze CA decarboxylation to taurine (Fig. 1i), but it was specific for CA, with CSA serving as a poor substrate (Fig. 1j and Extended Data Fig. 4c). The presence of CSA and its reaction product hypotaurine was inhibitory for CA decarboxylation (Extended Data Fig. 4d–f). To acknowledge its substrate specificity, we propose to name CA decarboxylase (CAD) the enzyme encoded by the *Gallus* gene annotated as CSAD based on its orthology[34]. We confirmed the opposite preference for the CSA substrate in human CSAD in our experimental conditions, and observed the same preference for CSA in a CSAD ortholog of a basal vertebrate (*Danio rerio*) (Fig. 1k). No conserved differences in residues of the active site cavity were found in *Gg*CAD and sauropsidian orthologs with respect to CSAD proteins from other vertebrates. However, two conserved substitutions (hydrophobic → hydrophilic) in sauropsidian sequences were observed in residues located within 5 Å from the active site cavity (Extended Data Fig. 5a). Analysis of the activity with the two substrates in single (Q467V and T470A) and double site-directed *Gg*CAD mutants, showed that these two substitutions contribute to the preference of the *Gallus* protein for CA (Extended Data Fig. 5b,c).

To determine the expression of CL, CBS and CSAD genes during early stages of embryogenesis, whole mount *in situ* hybridization analyses were performed in chicken embryos between 0.5 and 4 days of development (Hamburger-Hamilton [HH] stages 4–24)[35]. At HH stage 4, CL expression was first detected in the extraembryonic endoderm at the boundary of the area pellucida and area opaca (Fig. 2a). At HH stages 10 and 18, CL mRNAs were broadly detected throughout the extraembryonic endoderm (Fig. 2b,c). At

HH stage 18 and 24, widespread expression was also evident in the embryo proper (Fig. 2c,d). CBS expression was first detected at HH stage 4 weakly in the epiblast (Fig. 2e). At HH stage 10, CBS mRNAs were localized to the head region and in the intermediate mesoderm, with strong expression in the primitive blood cells of the extraembryonic blood islands (Fig. 2f). Broad CBS expression was evident throughout the embryo at HH stages 18 and 24 (Fig. 2g,h). CSAD expression was first detected at HH stage 4 in the extraembryonic endoderm (Fig. 2i). Expression in extraembryonic endoderm persisted at HH stages 10 and 18 (Fig. 2j,k). At HH stage 24, CSAD mRNAs were detected throughout the embryo, with higher levels of expression observed in liver and mesonephros (arrowhead and arrow, Fig. 2l). Inspection of the genomic regions containing CL, CBS, and CSAD genes revealed that CL is adjacent to CBS in a head-to-tail orientation on the chromosome 1 (Fig. 2m). Analysis of available RNA-seq profiles shows a prevalence of CBS over CL transcripts in the aggregated dataset. Tissue-specific RNA-seq profiles show abundant CBS transcripts in adult kidney and liver where CL transcripts are barely detected. Conversely, CL transcripts are more abundant than CBS transcripts in adult duodenum (Fig. 2m). CSAD is located on chromosome 33 adjacent to ZNF740 in a head-to-head orientation (Fig. 2n). The same organization is also observed for the human gene (www.ncbi.nlm.nih.gov/gene/51380), supporting orthology. CSAD transcripts are present in several adult tissues and especially abundant in kidney, liver, and duodenum (Fig. 2n).

By combining the evidence obtained with CL, CBS, and CAD proteins, one can define the pathway that produces sulfonated amino acids in embryonated chicken eggs (Fig. 3). The replacement of the thiol group of cysteine with sulfite by CL produces CA, which is decarboxylated by CAD to produce taurine (Fig. 3, upper branch). When these proteins are used together in the presence of sulfite, cysteine is quantitatively converted into taurine with little transient accumulation of CA (Extended Data Fig. 6a). Analysis of CL orthologs (see below) suggests that this pathway is universal in sauropsids. By contrast, the pathway is absent in synapsids or other vertebrates, in which taurine is formed through cysteine oxidation by cysteine dioxygenase (CDO) followed by CSA decarboxylation to hypotaurine and hypotaurine oxidation[19,22] (Fig. 3, lower branch). The sauropsidian pathway is shorter and does not involve formation of hypotaurine and sulfur oxidation. The H_2S produced by the CL reaction can be recycled for the formation of cysteine from other amino acids (Fig. 3, dashed line). In particular, the formation of cysteine from serine and H_2S is catalyzed by CBS (see Fig. 1g), whose gene is expressed at early stages in the chicken embryo (see Fig. 2e–h). CBS should be thus responsible for the serine hydrolase activity described in the chicken embryo liver[36]. By adding *in vitro Gg*CBS and serine to the other components of the pathway, a similar consumption of cysteine (Extended Data Fig. 6B) produces twice as much taurine (Extended Data Fig. 6c) with complete consumption of serine (Extended Data Fig. 6d).

Evolution of an enzyme able to catalyze the CL reaction has been key to the origin of the metabolic pathway. We found *bona fide* CL orthologs only in sauropsids, suggesting an origin of the protein family in this lineage. CL sequences form a separate group within the vertebrate CBS tree (Extended Data Fig. 7). In the maximum likelihood (ML) protein tree, the CL clade branches basal to teleostei (Extended Data Fig. 7a), while in the ML nucleotide tree, CL branches basal to amniotes (Extended Data Fig. 7b).

These phylogenetic reconstructions are complicated by differences in evolutionary rates and possible long-branch attraction artifacts[37] causing attraction of the fast evolving clade (CL) towards the basal clades. Rate differences between CL and CBS are mainly due to amino acid substitutions (Extended Data Fig. 7a,b). The ML tree obtained with the third (~synonymous) codon position showed reduced differences in branch lengths and the expected sister relationship of sauropsidian CBS and CL clades (Extended Data Fig. 7c). The CBS-CL locus (see Fig. 2m) is present in conserved synteny in all sauropsidian genomes and absent in non-sauropsids (Fig. 4). This suggests that CL evolved by tandem duplication of CBS after the split of sauropsids and synapsids in the late Paleozoic, c.a. 300 MYA. After duplication, one copy retained the original function while the other developed a novel catalytic ability through molecular changes involving one deletion and five substitutions of conserved active site residues; this neofunctionalization process was completed before separation of extant sauropsids (Fig. 4 and Extended Data Fig. 8). A further step involved adaptation of CSAD ortholog to the new function (co-option, Fig. 4) by promoting a secondary activity (CAD) to the main one. Interestingly, this tuning of substrate specificity occurred with substitution of residues located externally to the active site (see Extended Data Fig. 5). Based on evidence from extant genes and ancestral reconstruction, all sauropsids are expected to have inherited and maintained the CL pathway for taurine biosynthesis.

Discussion

Identification in this study of a route to taurine biosynthesis that originated in the common ancestor of birds and reptiles reveals that early amniote adaptation to terrestrial habitats -particularly embryonic development in the reptilian egg - also entailed the evolution of novel enzymes and metabolic pathways. Knowledge of the requirement of a particular cofactor (PLP) for the enzymatic activity has been key to the discovery of the CL gene since it allowed restriction in the search to a limited number of proteins expected to use that cofactor. The experimental characterization of the CL candidate provided clear-cut evidence for its functional assignment. The identification of the cysteic acid decarboxylase (CAD) as the enzyme responsible for the conversion of the CL reaction product into taurine has been a surprising outcome of the experimental validation process. No such enzyme has been previously described in literature. The *Gallus* ortholog of mammalian cysteine sulfinate decarboxylase (CSAD) was tested in view of the reported ability of CSAD to decarboxylate CA as secondary reaction. Unexpectedly, this protein revealed a strong preference for CA. The evolutionary shift of CSAD towards a preference for the CL reaction product (see Fig. 1k) gives independent support to the physiological relevance of the CL reaction and pathway for taurine biosynthesis. It further suggests that the CDO pathway (see Fig. 3) is less relevant for taurine production in sauropsids despite maintenance of a CDO gene in their genomes.

While physiological roles of taurine in specific organs such as retina, brain, muscles, and kidneys have been investigated only in mammals[19], there is evidence of the importance of taurine in the formation of bile in sauropsids[38–40]. Only taurine conjugates are found in chicken bile, owing to the inability of the *Gallus* bile acid-CoA:amino acid N-acyltransferase (BAAT) enzyme to use glycine in substitution of taurine[41]. During embryo development, bile aids utilization of lipids, the major source of energy of the egg yolk

and a source of metabolic water[42]. Fats are absorbed through the yolk sac membrane, by specialized endodermal cells containing bile and lipases[43]. Expression of the CL gene in the extraembryonic endoderm during early development (see Fig. 2a–d) suggests that formation of bile for fat digestion is a role of the taurine produced by this pathway. However, the more pleiotropic expression of the CSAD gene, especially during later development (see Fig. 2i–l), is consistent with a broader physiological role of taurine in different organs, such as regulation of cellular osmolarity[44] and cytoprotection[45], as observed in mammals.

Given that taurine biosynthesis is required in the reptilian egg, a question is why sauropsids evolved a different pathway. A possible advantage of the novel biosynthetic route is that it does not require molecular oxygen. Although gas exchange is ensured by microscopic pores in the egg shell, oxygen is limiting for the growth of chick embryos and critical for embryo survival in reptile species that nest underground[46,47]. The saving of reduced sulfur can be an additional advantage of the CL pathway, as oxidized sulfur (SO_3^{2-}) is used instead of cysteine oxidation to obtain sulfonated amino acids. In the chicken egg, also sulfate (SO_4^{2-}) is eventually incorporated into taurine by enzymatic conversion to sulfite[15]. Sulfite and sulfate are products of spontaneous oxidation of the sulfur present in the cell. Conversely, animals are unable to reduce oxidized sulfur to sulfide for its incorporation into amino acids. Therefore, in oviparous amniotes the reduced sulfur needed for embryo development until hatching must be stored in the egg. This sulfur content is witnessed by features observed in everyday use of unfertilized eggs: release of H_2S during heating contributes to the distinct egg flavour and is responsible for formation of FeS precipitates that turn green the yolk surface of hard-boiled eggs[48]. The CL pathway can provide a more efficient way to synthesize taurine by finding a use for oxidized sulfur that otherwise would be a waste -likely toxic[49] - product of cellular metabolism. In addition, functional and phylogenetic links between CBS and CL support the existence of a reduced sulfur cycle in the reptilian egg allowing the reuse of H_2S in amino acids (see Fig. 3). Origin and maintenance of this pathway in the saupsidian lineage provide evidence that the need to complete development out of water in a self-contained life-supporting structure imposes a selective pressure on embryo metabolism for efficient use of resources.

Methods

In silico subtraction of PLPomes

Gallus gallus and *Homo sapiens* PLPomes were compared side-by-side using the “whole genome analysis” tool of B6db (http://bioinformatics.unipr.it/B6db) with the option “exclude isoforms”. To facilitate identification of common and unique genes, the results were visually inspected with the help of the “highlight BRH” option on the results page painting in color entries that are Best Reciprocal Hits (BRH) in the two species. The comparison was extended to other sauropsids (e.g. *Anolis carolinensis*) and mammals (e.g. *Mus musculus*) to infer the conservation of unique genes in their respective taxonomic classes.

To determine active site conservation of the identified *Gallus* genes, substrate-binding cavities of human CBS (PDB: 4L3V) and CSAD (PDB: 2JIS) were determined through cavity computation by CAVER Analyst 2.0 BETA with Large Probe and Probe (respectively. 3.00 and 2.50 Å in CBS, and 4.00 and 2.80 in CSAD). Residues corresponding to the cavity

were highlighted in multiple alignments (see Extended Data Fig. 2A, 5A, and 8B) using Espript ver. 3.0 (http://espript.ibcp.fr).

Molecular phylogeny

Protein sequences were downloaded from NCBI and aligned with Clustalw 2.1. To obtain coding sequence (CDS) alignments, CDS were extracted from the corresponding mRNA sequences using ORFfinder 0.4.3 with the options "-s 0 -ml 1000 -strand plus -outfmt 1", and aligned based on amino acid alignment with macse v2.03. Phylogenetic trees were constructed with RAxML v. 7.7.8 using the general time reversible (GTR) amino acid substitution matrix with optimization of substitution rates and GAMMA model of rate heterogeneity. The partitioning of codon positions was specified in a partition file to generate separated alignment sets for the first and second codon positions and third codon position using the option '-f s'. Maximum-likelihood reconstruction of ancestral character states including insertions/deletions[51] was obtained with the FastML web server (http://fastml.tau.ac.il/) based on extant CBS and CL sequences and a phylogenetic tree assuming CBS duplication in the sauropsidian ancestor.

Embryo Collection and In Situ Hybridization

Fertile chicken eggs (Hy-Line, Iowa) were incubated in a humidified incubator at 37.5 °C for 0.5 to 5 days. Embryos were collected into chilled chick saline (123 mM NaCl), removed from the vitelline membrane and cleaned of yolk. Extra-embryonic membranes and large body cavities (brain vesicles, atria, allantois, eye) were opened to minimize trapping of the in situ reagents. Embryos were fixed overnight at 4°C in freshly prepared 4% paraformaldehyde in PBS, washed twice briefly in PBS plus 0.1% Triton X-100 then dehydrated through a graded MEOH series and stored at −20 °C overnight in 100% MEOH. cDNA templates for generating all antisense RNA probes were obtained by reverse transcriptase-polymerase chain reaction using pooled RNA from embryos between HH stages 4 and 30. Primer sequences were designed using the mRNA sequence in the NCBI database. Embryo processing, antisense RNA probe preparation and whole-mount ISHs were performed as described[52]. Experiments with fertilized eggs were conducted in accordance with federal agency guidelines. A detailed protocol is available for download at http://geisha.arizona.edu.

Plasmid Construction

For construction of *Gg*CL expression plasmids, the LOC418544 (NCBI GeneID: 418544) CDS sequence (XM_015300896) inserted into pcDNA3.1+/C-(K)DYK vector was purchased from GenScript (USA Inc.). The sequence was then amplified using CBSL_Fw, CBSL_Rev primers (native *Gg*CL) or CBSL_Fw, CBSL_short_Rev (truncated *Gg*CL) by PCR, using Phusion DNA polymerase, and inserted into pET-28b expression vector at NdeI/XhoI sites, generating respectively pET-28b-native*Gg*CL and pET-28b-truncated*Gg*CL. Details of the designed primers are reported in supplementary Table 1. A first transformation of the constructs into *E. coli* XL1Blue strain by electroporation was performed for plasmid amplification. Plasmids were extracted by alkaline lysis and transformed into *E. coli* BL21 Codon Plus strain by electroporation. For construction of *Gg*CAD expression plasmids, *Gg*CAD wild-type sequence (NCBI GeneID: 426184)

and mutated sequences (Q467V, T470A, Q467V-T470A) inserted into pET-28b expression vector were purchased from GenScript (USA Inc.), generating respectively pET-28b-*Gg*CAD, pET-28b-*Gg*CAD_Q467V, pET-28b-*Gg*CAD_T470A, pET-28b-*Gg*CAD_Q467V-T470A. The constructs were transformed directly into *E. coli* BL21 Codon Plus by electroporation. The authenticity of all constructs was verified by sequence analysis.

Protein expression and purification

Protein expression was performed inoculating a single colony of every clone in a Liter of autoinducing LB broth obtained by adding 0,5 g/L glucose and 2 g/L lactose to standard LB medium. Cells were grown at 30°C for 16h (*Gg*CL, *Gg*CBS), or at 20°C for 16h after a pre-induction phase at 30°C for 8h (*Gg*CAD). Cell pellets were resuspended in 50 mL of Lysis Buffer (NaH_2PO_4 20 mM pH 7.0, NaCl 100 mM, 20 μM PLP, 50 mM imidazole), sonicated (1s on/off alternatively at 30 W for 30 min) and centrifuged (14000 rpm for 40 minutes). Supernatant was loaded onto a 50 mL Superloop of AKTA pure system FPLC and purified by Affinity Chromatography (AC) using HisTrap 5 mL FF column. Proteins were eluted with AC Elution Buffer (NaH_2PO_4 20 mM pH 7.0, NaCl 100 mM, 20 μM PLP, 500 mM imidazole). *Gg*CL fractions were collected and diluted in 50 mL of Loading Buffer (MES 20 mM pH 6.5, 20 μM PLP) for Cation Exchange Chromatography (CIEX) using HiTrap SP FF column, and eluted in 35 mL gradient of CIEX Elution Buffer (MES 20 mM pH 6.5, 1M NaCl, 20 μM PLP). Protein fractions (see Extended Data Fig. 2b) were collected and concentrated by Vivaspin™ centrifugation for further purification steps. Size Exclusion Chromatography (SEC) was performed with Superdex 200 column using SEC Buffer (NaH_2PO_4 20 mM pH 7.0, NaCl 100 mM) used as well as Storage Buffer. *Gg*CAD fractions after AC were collected and diluted in 50 mL of Loading Buffer (NaH_2PO_4 20 mM pH 8.0, 20 μM PLP) for Anion Exchange Chromatography (AIEX) using HiTrap Q FF column, and eluted in 35 mL gradient of AIEX Elution Buffer (NaH_2PO_4 20 mM pH 8.0, 1M NaCl, 20 μM PLP).

UV-Visible and fluorescence spectroscopy

JASCO spectrophotometer was used to measure absorbance spectra of purified enzymes and for determination of kinetic parameters of *Gg*CL. Eluted fractions of enzymes were measured for protein quantification. For the quantification of *Gg*CL, absorbance at 428 nm (Soret peak) was measured, using an extinction coefficient of 84900 M^{-1} cm^{-1} previously determined for *Hs*CBS[53]. For the quantification of different CSAD/CAD proteins, absorbance at 280 nm was measured, using molar extinction coefficients computed with ProtParam (57410 M^{-1} cm^{-1} *Gg*CAD, 61880 M^{-1} cm^{-1} *Hs*CSAD, 72880 M^{-1} cm^{-1} *Dr*CSAD).

H_2S release due to CL reactions was followed spectrophotometrically at 390 nm as formation of PbS using the previously calculated 54 extinction coefficient of 5500 M-1 cm-1. Reaction mixture were prepared in a 1 mL plastic cuvette with 50 mM NaH_2PO_4 pH 7.0, different concentrations of cysteine and sulfite, and 0.4 mM of lead acetate; the reaction was started by addition of 1 μM *Gg*CL. Velocity with different concentrations of cysteine and sulfite were taken at maximum speed reached by the enzyme, that did not correspond to

the initial velocity of the kinetic, due to an appreciable delay at start (see Extended Data Fig. 2g). Data were fitted to the Michaelis-Menten equation using SigmaPlot 14.0.

The presence of PLP bound to the protein was assessed by fluorescence spectroscopy[55]. Fluorescence measurements were performed on a FluoroMax-3 spectrofluorometer (HORIBA Jobin Yvon, Kyoto, Japan) equipped with a thermostatic bath, set at 20 °C. *Gg*CL concentration was 40 μM in 20 mM NaH_2PO_4 pH 7.0, 100 mM NaCl; *Gg*CAD concentration was 20 μM in 20 mM NaH_2PO_4 pH 8.0, 100 mM NaCl. Excitation and emission slits width was set to 7 nm and the integration time to 0.6 seconds. *Gg*CL emission spectrum was recorded between 425 nm and 600 nm using an excitation wavelength of 412 nm. *Gg*CAD emission spectrum was recorded between 355 nm and 500 nm using an excitation wavelength of 340 nm. The spectra were corrected for buffer contribution.

NMR Spectroscopy

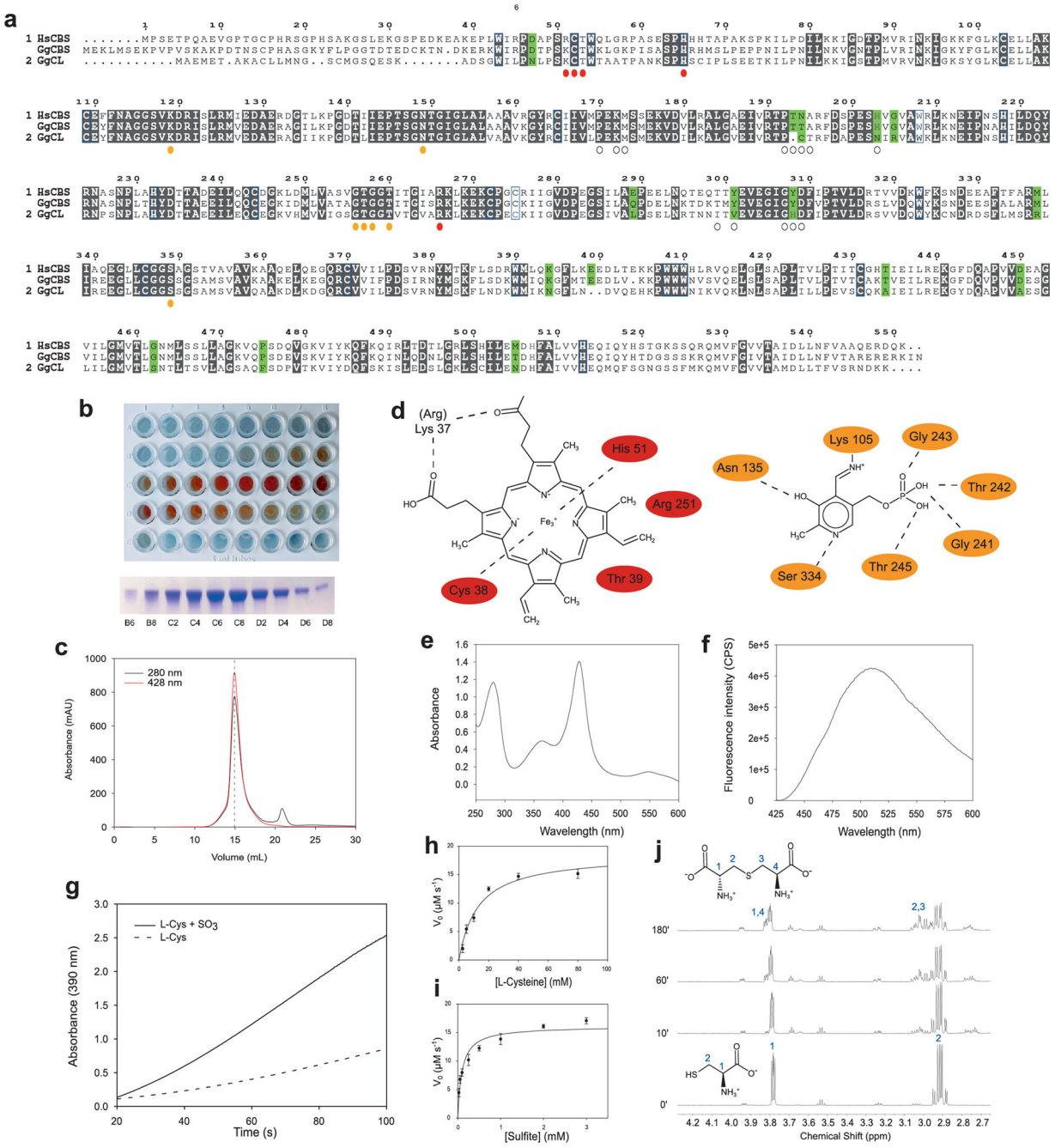
^1H NMR spectra were acquired with a JEOL ECZ600R spectrometer in no spinning mode. Samples were loaded in Wilmad ECONOMY NMR tubes, solved in 600 ul of H_2O:D_2O (9:1). For single spectra measurement (i.e. substrates spectra t_0) we used simple DANTE presat sequence for H_2O suppression. To monitor the reaction kinetics, we use a kinetic array of DANTE presat sequence with 600 min periods for 1h, 2h, 3h depending on reaction speed. NMR experiments were performed with 50 mM NaH_2PO_4 pH 7.0 to avoid signals of organic buffers in ^1H NMR spectra. The time-courses of the CAD reaction at a given initial substrate concentration were fitted with the integrated Michaelis-Menten equation[50] using the R software. The R script containing the equation and commands (SM.tar.gz) used to produce the fitting shown in Fig 1j is provided in the dataset deposited at the Harvard dataverse repository (https://doi.org/10.7910/DVN/UYAUBO). Specific activities of CSAD/CAD enzymes and mutants were determined by quantifying the reaction products obtained with CSA and CA substrates after 5 minutes of reactions stopped with 1 M HCl. All the spectra and kinetics were collected at 25°C.

Extended Data

		Gallus gallus		*Homo sapiens*		
Activity	EC (subfamily)	Accession	E-value	Accession	E-value	
Glycine dehydrogenase (decarboxylating).	1:4:4:2	XP_015135548.1	0.0E+000	NP_000161.2	0	
Glycine hydroxymethyltransferase.	2:1:2:1	XP_414824.4	1.6E-287	NP_004160.3	4.2E-295	
				NP_001159828.1	2.4E-277	
Glycine C-acetyltransferase.	2:3:1:29	XP_004950962.1	4.4E-227	NP_001165161.1	3.7E-225	
5-aminolevulinic acid synthase.	2:3:1:37	XP_025010275.1	1.6E-270	NP_000023.2	1.7E-269	
				NP_954635.1	1E-268	
Serine C-palmitoyltransferase.	2:3:1:50 (a)	XP_004949309.1	4.3E-231	NP_006406.1	2.2E-237	
	2:3:1:50 (b)	NP_001006483.1	1.7E-294	NP_004854.1	7.4E-299	
	2:3:1:50 (b)	XP_025005211.1	3.5E-275	NP_001336874.1	4.1E-260	
Phosphorylase.	2:4:1:1	NP_001026205.1	0.0E+000	NP_002853.2	0	
	2:4:1:1	NP_989723.1	0.0E+000	NP_005600.1	0	
				NP_002854.3	0	
Aspartate aminotransferase.	2:6:1:1 (a)	NP_990854.1	1.4E-297	NP_002070.1	5.7E-296	
	2:6:1:1 (a)	NP_990652.1	5.2E-297	NP_002071.2	1.2E-305	
Ornithine--oxo-acid aminotransferase.	2:6:1:13	NP_001006567.1	7.9E-276	NP_001309897.1	3.4E-275	
4-aminobutyrate aminotransferase.	2:6:1:19 (a)	XP_414940.2	0.0E+000	XP_011520702.1	0	
2-aminoadipate aminotransferase.	2:6:1:39 (a)	XP_426286.3	3.6E-236	XP_006714294.1	4.1E-265	
Branched-chain amino acid aminotransferase.	2:6:1:42	XP_416424.1	3.0E-184	NP_005495.2	3.3E-191	
				XP_024307398.1	1.2E-181	
Alanine--glyoxylate aminotransferase.	2:6:1:44	XP_429219.3	0.0E+000	XP_005248394.1	0	
Tyrosine aminotransferase.	2:6:1:5	XP_025010115.1	3.1E-253	NP_000344.1	5.8E-267	
Serine--pyruvate aminotransferase.	2:6:1:51	XP_003641783.2	2.9E-242	NP_000021.1	1.1E-259	
Phosphoserine aminotransferase.	2:6:1:52	XP_424846.3	2.8E-232	NP_478059.1	1E-234	
Kynurenine--oxoglutarate aminotransferase.	2:6:1:7	XP_415485.2	5.8E-262	XP_016870749.1	1.5E-278	
	2:6:1:7	XP_025009029.1	4.8E-268	NP_001336377.1	8.9E-268	
Alanine aminotransferase.	2:6:1:2 (b)			NP_005300.1	6.4E-262	
		XP_015147909.1	9.8E-259	NP_597700.1	4E-267	
Cysteine desulfurase.	2:8:1:7 (a)	NP_001026018.1	3.4E-297	NP_066923.3	1.9E-298	
molybdenum cofactor sulfurtransferase	2:8:1:9	XP_419048.4	1.9E-290	NP_060417.3	0	
O-phospho-L-seryl-tRNA(Sec):L-selenocysteinyl-tRNA synthase	2:9:1:2	NP_001026329.1	0.0E+000	NP_058651.3	0	
Kynureninase.	3:7:1:3	XP_004943069.1	9.1E-240	XP_024308976.1	3.1E-259	
Glutamate decarboxylase.	4:1:1:15 (a)	NP_990244.1	0.0E+000	NP_000808.2	0	
	4:1:1:15 (a)	XP_015137540.1	0.0E+000	NP_001127838.1	0	
Ornithine decarboxylase.	4:1:1:17:1	NP_001161238.1	1.7E-171	NP_002530.1	5.2E-174	
Histidine decarboxylase.	4:1:1:22 (b)	NP_001280217.1	2.1E-283	NP_002103.2	1.3E-285	
Aromatic-L-amino-acid decarboxylase.	4:1:1:28	XP_419032.3	3.7E-275	NP_001076440.1	1.9E-285	
Sulfinoalanine decarboxylase.	4:1:1:29	XP_004939481.1	0.0E+000	NP_997242.2	0	
	4:1:1:29	XP_025001257.1	1.0E-283	NP_057073.4	0	
Sphinganine-1-phosphate aldolase.	4:1:2:27	NP_001007947.1	3.5E-255	NP_003892.2	1.1E-248	
Threonine aldolase.	4:1:2:5	XP_015151050.1	1.0E-164			**probable low-specificity L-threonine aldolase 2**
Cystathionine beta-synthase.	4:2:1:22	XP_015156364.1	7.4E-231	NP_001171479.1	7.5E-228	
	4:2:1:22	XP_015156382.1	4.8E-214			**cystathionine beta-synthase-like**
Threonine synthase.	4:2:3:1	XP_025005939.1	6.4E-083	NP_060741.3	2.8E-080	
	4:2:3:1	XP_025004138.1	1.2E-093	NP_079114.3	1.2E-090	
Ethanolamine-phosphate phospho-lyase	4.2.3.2	XP_015132066.2	1.6E-119	NP_112569.2	1.3E-114	
5-phosphonooxy-L-lysine phospho-lyase	4.2.3.134	XP_025010761.1	1.0E-111	NP_699204.1	9.3E-100	
L-serine ammonia-lyase.	4:3:1:17a	NP_001185572.1	5.4E-204	NP_612441.1	1.8E-205	
				NP_006834.2	8.4E-209	
D-serine ammonia-lyase.	4:3:1:18 (b)	XP_003640696.1	7.2E-172			**uncharacterized protein CHDSD**
Cystathionine gamma-lyase.	4:4:1:1	XP_422542.2	2.2E-280	NP_001893.2	5.1E-291	
1-aminocyclopropane-1-carboxylate synthase.	4:4:1:14	XP_015142682.1	3.2E-082	NP_001027025.2	3.4E-078	
	4:4:1:14	XP_015142683.1	4.6E-081	NP_115981.1	1.4E-081	
Selenocysteine lyase.	4:4:1:16 (b)	NP_001132935.1	0.0E+000	NP_057594.4	0	
Serine racemase.	5:1:1:18	XP_001234891.1	6.6E-180	XP_006721628.1	1.4E-223	
Ornithine decarboxylase paralogue		NP_001280585.1	8.9E-145	NP_443724.1	1E-196	
		NP_001008729.1	4.2E-110	NP_056962.2	2.3E-116	
L-phenylserine dehydratase		XP_004942032.2	2.9E-049			**uncharacterized protein SRRL**
GOT1L1				NP_689626.2	0	
PROSC		XP_424381.4	2.4E-132	NP_009129.1	5.2E-150	
PDXDC1		XP_004945124.1	9.3E-007	NP_001272379.1	0.00000052	
PDXDC2				XP_016885696.1	0.00000011	

Extended Data Fig. 1. *In silico* subtraction of chicken and human PLPomes.

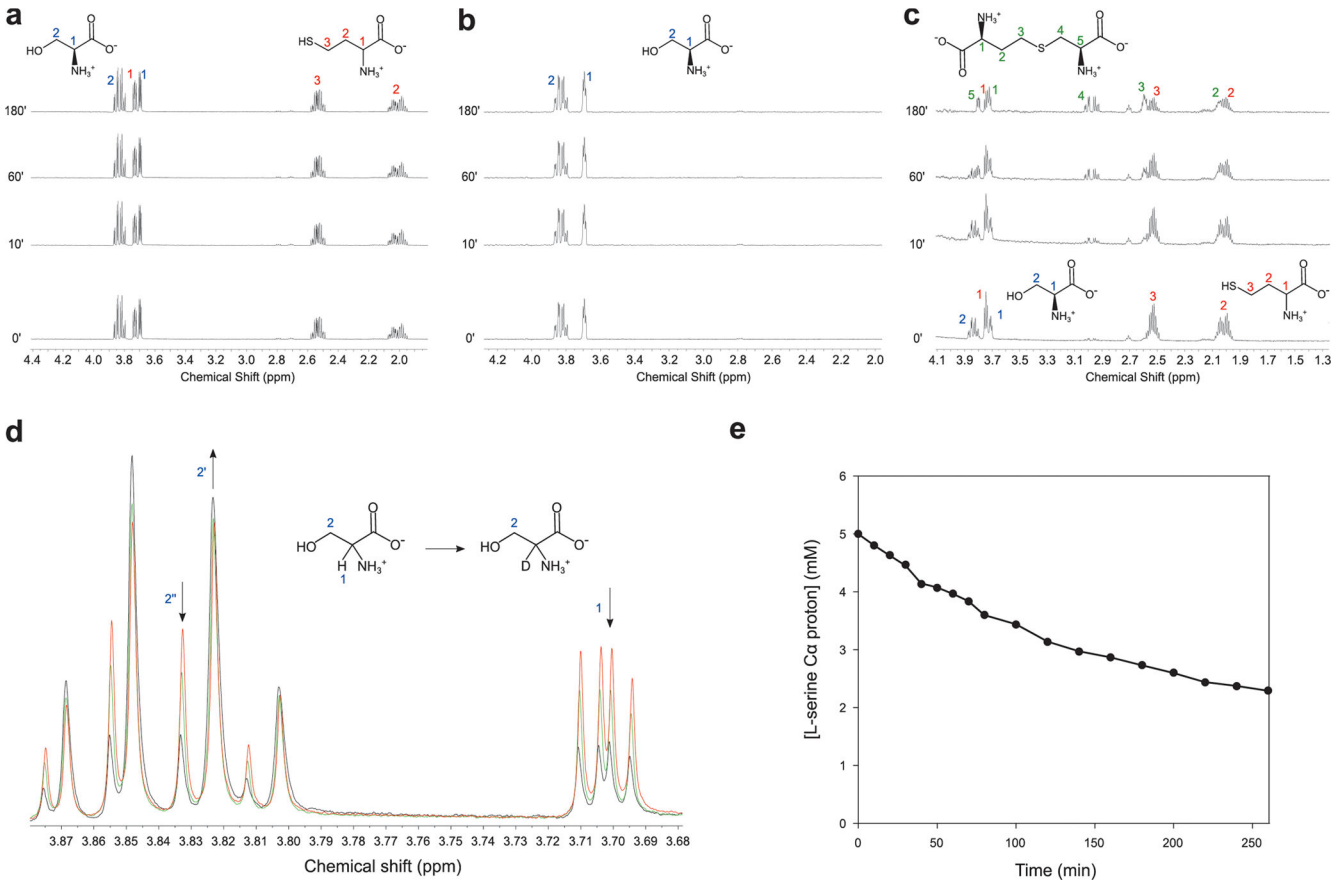
Comparison of the complete set of PLP-dependent enzymes (one isoform per gene) in *Gallus gallus* and *Homo sapiens* as classified by B6db. Orthologous proteins (BRH test) are colored blue. *Gallus* proteins without human orthologs are in bold. E-values indicate significance of the protein alignments with family-level Hidden Markov Models.



Extended Data Fig. 2. *Gg*CL is a heme and PLP protein with cysteine lyase activity.

a, Multiple alignment of *H. sapiens* CBS (*Hs*CBS) with *G. gallus* CBS and CL proteins (*Gg*CBS, *Gg*CL). Filled circles indicate residues that recognize heme (red), PLP (yellow) and serine (white) in the holo CBS structure (PDB code 3PC4). Conserved residues based on the alignment of 8 CL and 22 CBS sequences from vertebrates are shaded in black. Green shading indicates conserved differences between CBS and CL groups. **b**, Photograph of the FPLC collector after cation exchange, showing the vivid orange color of *Gg*CL protein fractions (upper panel); selected fractions were subjected to SDS-PAGE electrophoresis and stained with Coomassie Brilliant Blue (lower panel). **c**, Gel filtration chromatogram

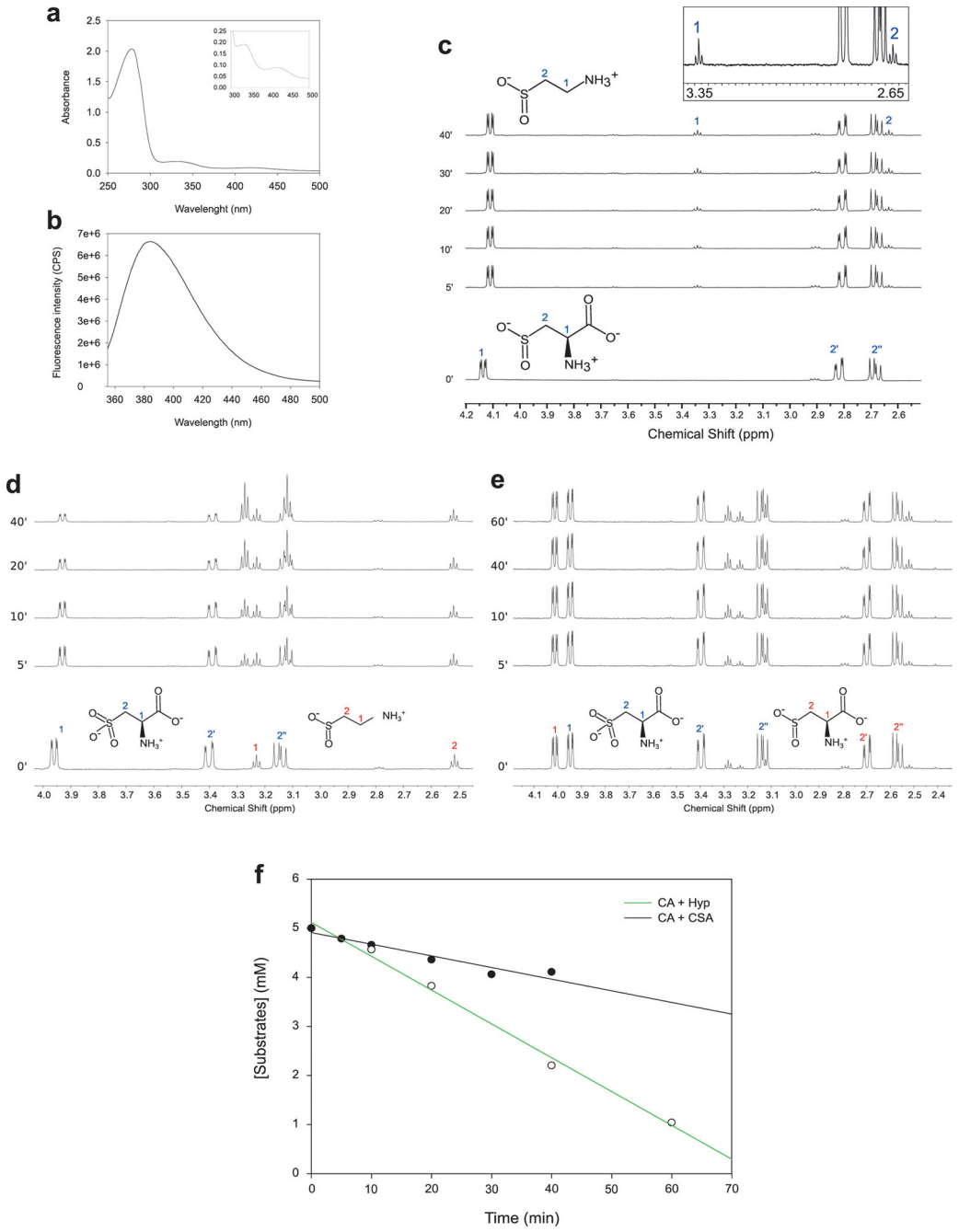
(Superdex 200) with dual wavelength detection (λ = 280, 428 nm), showing a molecular weight corresponding to *Gg*CL monomer. **d**, *Gg*CL predicted interactions with heme (left) and PLP (right) are shown with residues conserved in the alignment of CBS/CL proteins highlighted in colors. **e**, Absorbance spectrum of recombinant *Gg*CL (16.5 μM) in NaH_2PO_4 (20 mM), pH 7.0. **f**, Fluorescence emission spectrum (excitation: 412 nm) of recombinant *Gg*CL (22 μM) in NaH_2PO_4, pH 7.0. **g**, Kinetics of H_2S release by the CL reaction monitored spectrophotometrically at 390 nm in 50 mM NaH_2PO_4, pH 7.0 with *Gg*CL (1 μM), lead acetate (0.4 mM), cysteine (5 mM) in the absence (dashed line) or in the presence of Na_2SO_3 (5 mM, solid line). **h-i**, Non linear fitting to the Michaelis Menten equation of the dependency on substrate concentrations of the initial reaction velocity of *Gg*CL (1 μM) with fixed (h) Na_2SO_3 (5 mM) and (i) cysteine (40 mM). Data are means ± SDV of three independent experiments. **j**, Time-resolved ^1H NMR spectra of cysteine (10 mM) in the presence of *Gg*CL (1 μM), showing partial conversion into lanthionine.



Extended Data Fig. 3. Absence of CBS activity in *Gg*CL.

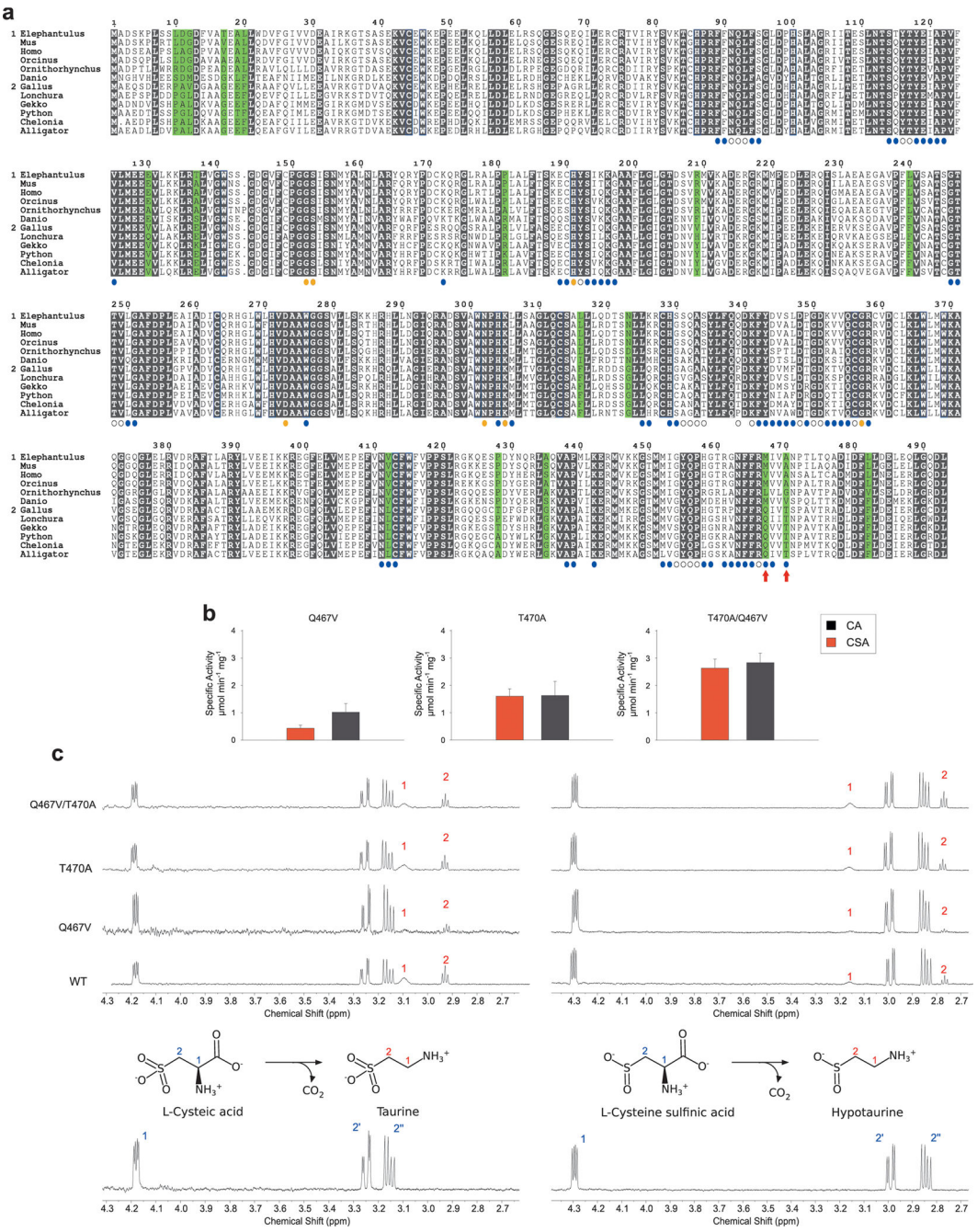
a, Time-resolved ^1H NMR spectra of 5 mM of serine (atoms labeled in blue) and 5 mM of DL-homocysteine (atoms labeled in red) in the presence of *Gg*CL (1 μM). **b**, Time-resolved ^1H NMR spectra of serine (5 mM) and Na_2S (5 mM) in the presence of *Gg*CL (1 μM). **c,** Time-resolved ^1H NMR spectra of 5 mM of serine (atoms labeled in blue) and 10 mM of DL-homocysteine (atoms labeled in red) in the presence of *Gg*CBS (4 μM), showing complete consumption of serine and partial conversion of DL-homocysteine

in cystathionine (atoms labeled in green) due to the stereospecific enzymatic reaction. **d**, Hydrogen-Deuterium exchange of serine alpha proton catalysed by *Gg*CL (1 μM) in 95% D_2O. Spectra were superimposed at time 0' (red), 60' (green), 260' (black). **e**, ^1H peak integration of serine Cα proton is plotted in the interval 0'−260'.



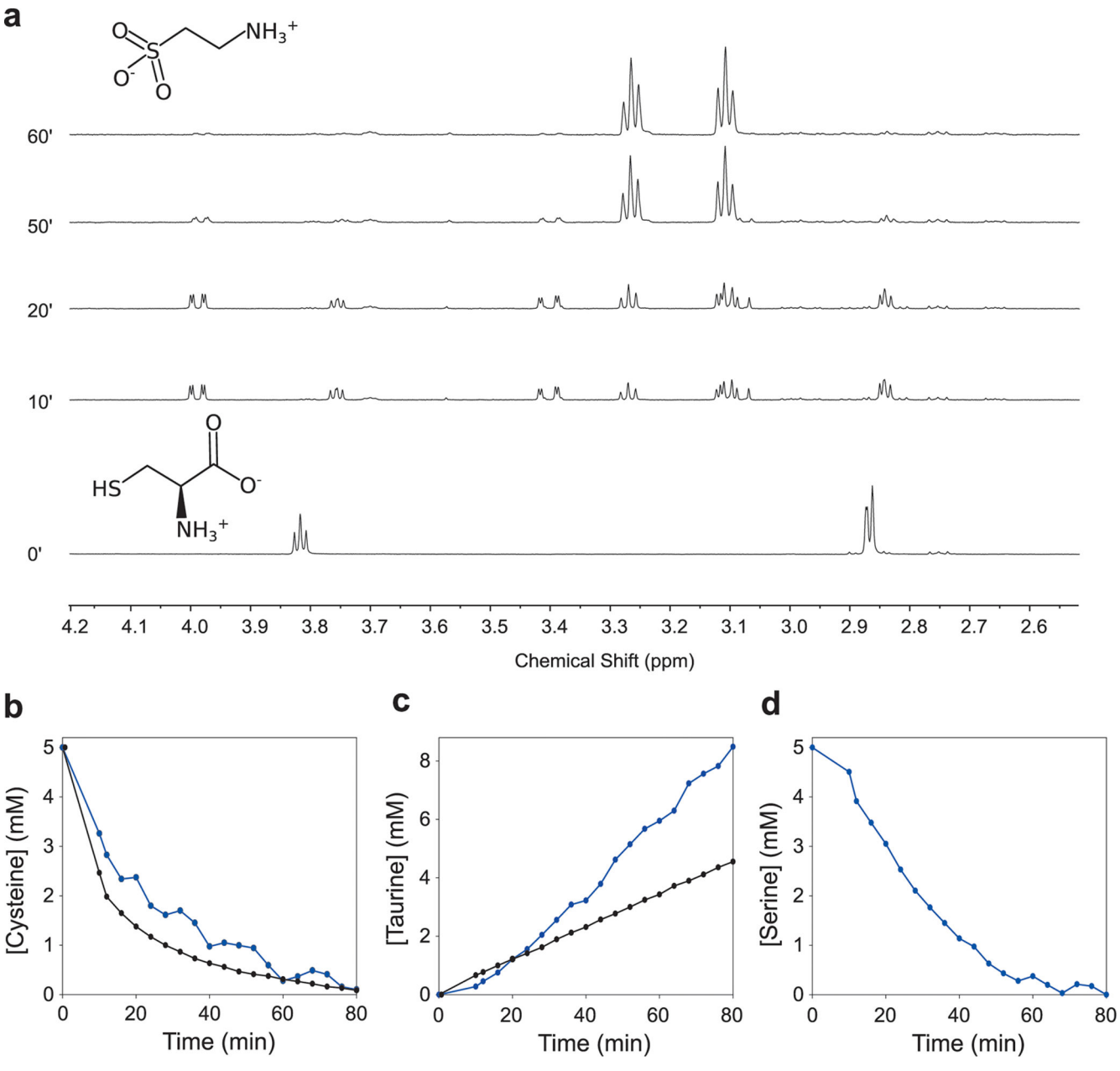
Extended Data Fig. 4. *Gallus* CSAD encodes a PLP-dependent cysteic acid decarboxylase (CAD). **a**, Absorbance spectrum of *Gg*CAD in 20 mM NaH_2PO_4, pH 8.0 and 100 mM NaCl; The absorbance region of PLP tautomers (enolimine 340 nm, ketoenamine 415 nm) is shown in the inset. b, Fluorescence emission spectrum of PLP enolimine tautomer upon excitation at

340 nm. **c**, Time-resolved ^{1}H NMR spectra of cysteine sulfinic acid (5 mM) in the presence of *Gg*CAD (1 μM), showing partial formation of hypotaurine (inset). **d**, Time-resolved ^{1}H NMR spectra of 5 mM of cysteic acid (atoms labeled in blue) and 5 mM of hypotaurine (atoms labeled in red) in the presence of *Gg*CAD (1 μM), showing slight inhibition of CAD activity. **e**, Time-resolved ^{1}H NMR spectra of 5 mM of cysteic acid (atoms labeled in blue) and 5 mM of cysteine sulfinic acid (atoms labeled in red) in the presence of *Gg*CAD (1 μM), showing strong inhibition of CAD activity. **f**, ^{1}H peak integration of CA signals in the presence of *Gg*CAD and hypotaurine (CA + Hyp) or cysteine sulfinic acid (CA + CSA).



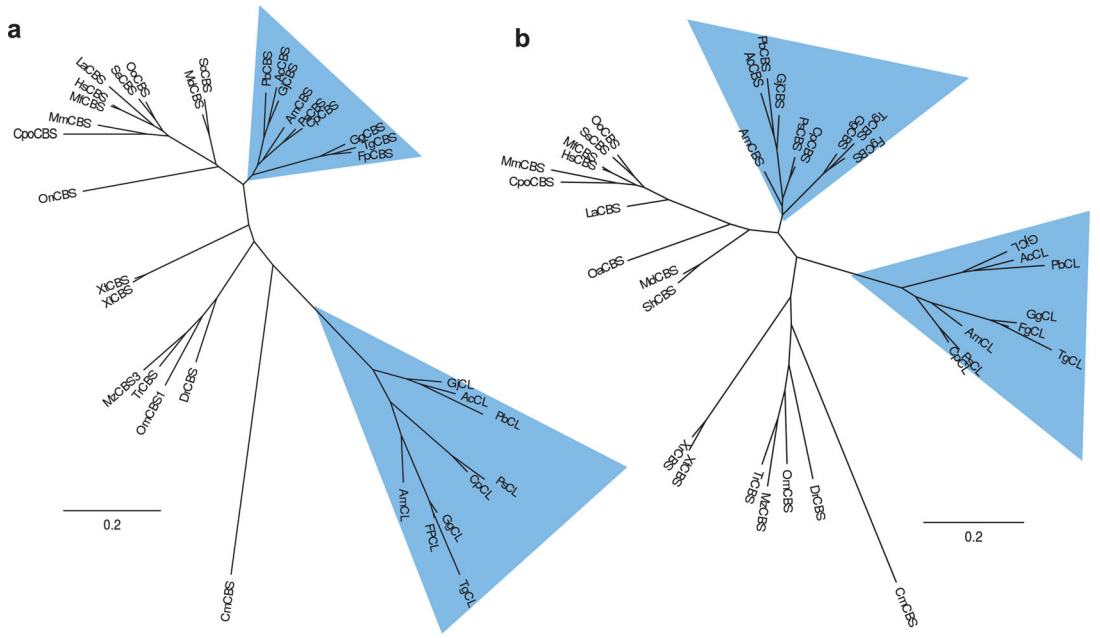
Extended Data Fig. 5. Analysis of *Gallus* CSAD site-directed mutants.

a, Multiple alignment of CSAD orthologs from (1) non-sauropsids and (2) sauropsids. Conserved differences between groups are shaded in green. Residues that recognize PLP (yellow) or line the active site cavity (white) or are within 5 Å from the active site cavity (blue) in the human holo CSAD structure (PDB code 2JIS) are indicated by filled circles; positions of site-directed mutants are indicated by red arrows. **b**, Specific activities of wild-type (WT), single (Q467V, T470A), and double (Q467V/T470A) *Gg*CAD mutants with CA and CSA substrates. **c**, ^{1}H NMR spectra showing decarboxylation activity of wild-type (WT), single (Q467V, T470A) and double (Q467V-T470A) mutants in the presence of cysteic acid (right) and cysteine sulfinic acid (left) after 5’ of reaction stopped with 1M HCl.



Extended Data Fig. 6. One-pot enzymatic synthesis of taurine from cysteine.

a, Time-resolved ^{1}H NMR spectra of cysteine (5 mM) and sulfite (7 mM) in the presence of recombinant *Gg*CL (1 μM) and *Gg*CAD (1 μM) proteins. **b-d**, ^{1}H peak integration of (b) cysteine, (c) taurine, and (d) serine NMR signals in the same reaction conditions as (a) in the absence (black dots) or in the presence (blue dots) of serine (5 mM) and *Gg*CBS (4 μM).

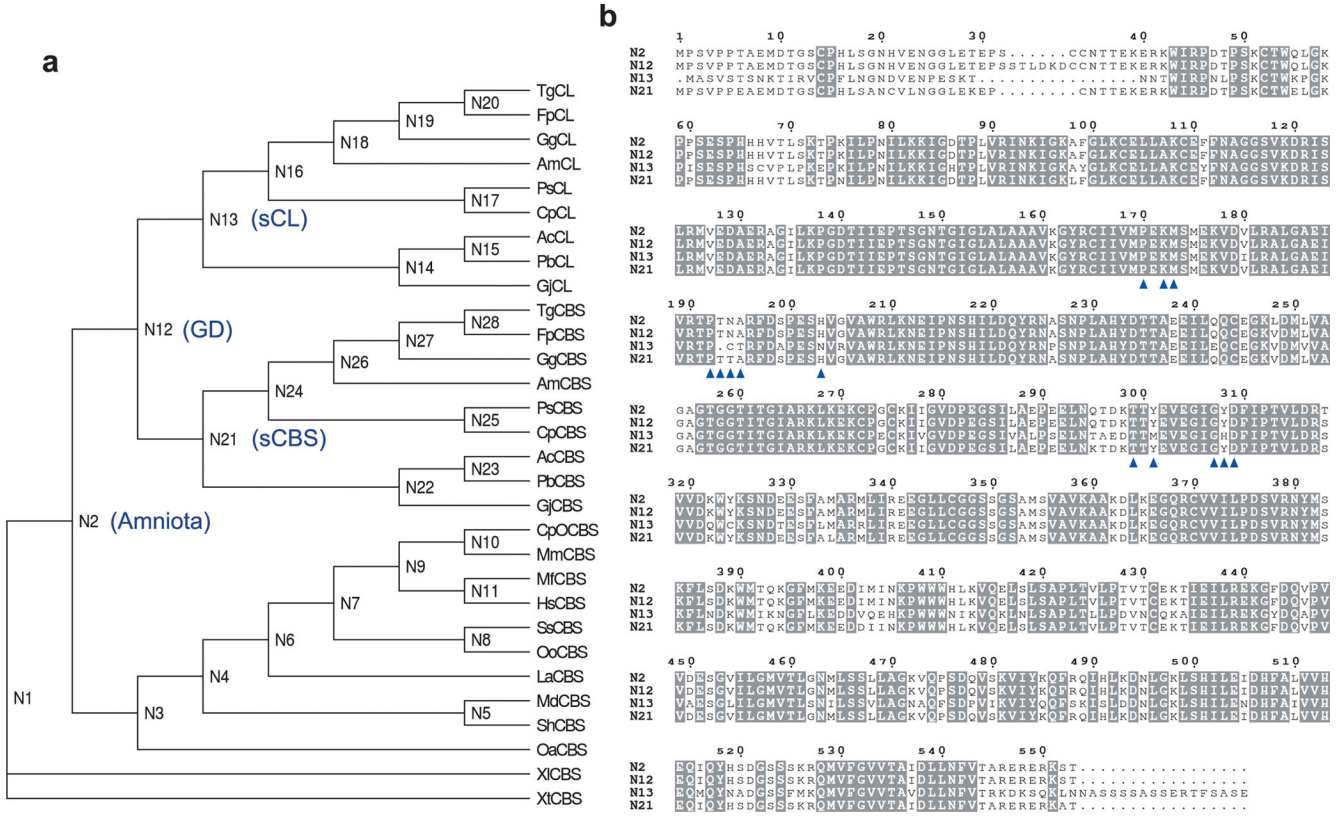


Name	Protein id	mRNA id	Taxonomy	Species
GgCBS	XP_416752.3	XM_416752.5	Sauropsida	*Gallus gallus*
GgCL	XP_015156382.1	XM_015300896.1	Sauropsida	*Gallus gallus*
TgCBS	XP_002189658.2	XM_002189622.3	Sauropsida	*Taeniopygia guttata*
TgCL	XP_012424887.1	XM_012569433.1	Sauropsida	*Taeniopygia guttata*
FpCBS	XP_027641626.1	XM_027785825.1	Sauropsida	*Falco peregrinus*
FpCL	XP_013154524.1	XM_013299070.2	Sauropsida	*Falco peregrinus*
AmCBS	XP_006260536.1	XM_006260474.3	Sauropsida	*Alligator mississippiensis*
AmCL	XP_019342804.1	XM_019487259.1	Sauropsida	*Alligator mississippiensis*
CpCBS	XP_005311024.1	XM_005310967.3	Sauropsida	*Chrysemys picta bellii*
CpCL	XP_023956308.1	XM_024100540.1	Sauropsida	*Chrysemys picta bellii*
PsCBS	XP_014436483.1	XM_014580997.1	Sauropsida	*Pelodiscus sinensis*
PsCL	XP_006137941.1	XM_006137879.2	Sauropsida	*Pelodiscus sinensis*
AcCBS	XP_003218996.1	XM_003218948.3	Sauropsida	*Anolis carolinensis*
AcCL	XP_008105670.1	XM_008107463.1	Sauropsida	*Anolis carolinensis*
PbCBS	XP_007435006.1	XM_007434944.2	Sauropsida	*Python bivittatus*
PbCL	XP_015744715.1	XM_015889229.2	Sauropsida	*Python bivittatus*
GjCBS	XP_015265299.1	XM_015409813.1	Sauropsida	*Gekko japonicus*
GjCL	XP_015265251.1	XM_015409765.1	Sauropsida	*Gekko japonicus*
HsCBS	NP_000062.1	NM_000071.2	Mammals	*Homo sapiens*
MfCBS	NP_001270081.1	NM_001283152.1	Mammals	*Macaca fascicularis*
MmCBS	NP_835742.1	NM_178224.3	Mammals	*Mus musculus*
CpoCBS	XP_003463971.2	XM_003463923.4	Mammals	*Cavia porcellus*
OoCBS	XP_004264645.1	XM_004264597.2	Mammals	*Orcinus orca*
SsCBS	XP_020926705.1	XM_021071046.1	Mammals	*Sus scrofa*
LaCBS	XP_010596971.1	XM_010598669.2	Mammals	*Loxodonta africana*
ShCBS	XP_012399820.1	XM_012544366.2	Mammals	*Sarcophilus harrisii*
MdCBS	XP_007493261.1	XM_007493199.2	Mammals	*Monodelphis domestica*
OaCBS	XP_007665408.2	XM_007667218.2	Mammals	*Ornithorhynchus anatinus*
XlCBS	NP_001086820.1	NM_001093351.1	Amphibia	*Xenopus laevis*
XtCBS	NP_001008171.1	NM_001008170.1	Amphibia	*Xenopus tropicalis*
DrCBS	NP_001104702.1	NM_001111232.1	Teleostei	*Danio rerio*
TrCBS	XP_003966607.1	XM_003966558.2	Teleostei	*Takifugu rubripes*
MzCBS	XP_004570219.1	XM_004570162.5	Teleostei	*Maylandia zebra*
OmCBS	XP_021464144.1	XM_021608469.1	Teleostei	*Oncorhynchus mykiss*
CmCBS	XP_007906994.1	XM_007908803.1	Chondrichthyes	*Callorhinchus milii*

Extended Data Fig. 7. Phylogeny of CBS and CL proteins in vertebrates.

Unrooted maximum-likelihood (ML) trees obtained from protein and nucleotide alignments of 35 CBS and CL sequences from 26 vertebrate species. Protein and nucleotide accession

numbers corresponding to tree tip names are indicated; sauropsidian sequences are shaded in blue. Scale bars represent the number of calculated substitutions per site. **a**, Protein ML tree (436 alignment patterns) showing branching of the CL clade basal to teleostei. **b**, Nucleotide ML tree (1277 alignment patterns) showing branching of the CL clade basal to amniotes. **c**, Third codon position ML tree (613 alignment patterns) showing branching of the CL clade within sauropsida.



c

	193		194		195		203		301		308	
N2 (Amniota)	T	0.99936	N	0.54534	A	0.99914	H	0.99992	Y	0.99988	Y	0.99992
N12 (GD)	T	0.99934	N	0.59071	A	0.99910	H	0.99990	Y	0.99986	Y	0.99990
N13 (sCL)	-	0.99000	C	0.99937	T	0.98614	N	0.99680	M	0.99757	H	0.99777
N21 (sCBS)	T	0.99998	T	0.99711	A	0.99996	H	0.99998	Y	1.00000	Y	1.00000

Extended Data Fig. 8. Ancestral substitutions in CL neofunctionalization.

a, Evolutionary dendrogram used in ancestral state reconstructions assuming split of amniote last common ancestor (Amniote; N2) into two lineages before the gene duplication (GD; N12) leading to saurospidian CL (sCL; N13) and CBS (sCBS; N21). Sequence identifiers are as in Figure S7. **b**, Multiple alignment of reconstructed ancestral sequences corresponding to nodes N2, N12, N13, and N21. Active site residues are indicated by blue triangles. Positions with Identical residues in the four nodes and human CBS are shaded gray. Numeration is in accordance with the human CBS sequence. **c**, Character state probabilities for active site residues substituted in *Gg*CL showing high probability of fixation before the split of extant sauropsids.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Davide Cavazzini, Roberto Tirindelli, and Angelo Totaro for help and discussion. This work was supported by the Italian Ministry for Education, University and Research (MIUR) PRIN 2017 grant 2017483NH8_003 to RP and benefited from the equipment and framework of the COMP-HUB initiative, supported by the MIUR 'Departments of Excellence' program 2018–2022 and from the High Performance Computing (HPC) facility of the University of Parma, Italy. The chicken in situ hybridization studies were supported by NIH grant P41HD088362 to PBA.

Data availability

Data and supplementary information are available in the manuscript. Raw data, computer code, sequence alignments and trees for Figs 1–4 and Extended Data Figs are deposited in the Harvard dataverse repository https://doi.org/10.7910/DVN/UYAUBO. PLPomes of the amniotes analyzed in this study and other organisms can be accessed and compared with the B6 database http://bioinformatics.unipr.it/B6db.

Bibliography

1. Kumar S, Stecher G, Suleski M & Hedges SB Timetree: A resource for timelines, timetrees, and divergence times. Mol. Biol. Evol 34, 1812–1819 (2017). [PubMed: 28387841]
2. Reisz RR & Müller J Molecular timescales and the fossil record: a paleontological perspective. Trends Genet 20, 237–241 (2004). [PubMed: 15109777]
3. Ford DP & Benson RBJ The phylogeny of early amniotes and the affinities of Parareptilia and Varanopidae. Nat. Ecol. Evol 4, 57–65 (2020). [PubMed: 31900445]
4. Sander PMPaleontology. Reproduction in early amniotes. Science337, 806–808 (2012). [PubMed: 22904001]
5. Pough FH, Janis CM & Heiser JB Vertebrate life (Pearson, 2013).
6. Wright PANitrogen excretion: three end products, many physiological roles. J. Exp. Biol198, 273–281 (1995). [PubMed: 7699310]
7. Packard GCThe influence of ambient temperature and aridity on modes of reproduction and excretion of amniote vertebrates. Am. Nat100, 667–682 (1966).
8. Salway JGThe krebs uric acid cycle: A forgotten krebs cycle. Trends Biochem. Sci43, 847–849 (2018). [PubMed: 29807701]
9. Mommsen TP & Walsh PJ Evolution of urea synthesis in vertebrates: the piscine connection. Science 243, 72–75 (1989). [PubMed: 2563172]
10. Bairoch AThe ENZYME database in 2000. Nucleic Acids Res28, 304–305 (2000). [PubMed: 10592255]
11. Kanehisa M, Sato Y, Kawashima M, Furumichi M & Tanabe M KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44, D457–62 (2016). [PubMed: 26476454]
12. Caspi Ret al.The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res46, D633–D639 (2018). [PubMed: 29059334]
13. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature432, 695–716 (2004). [PubMed: 15592404]
14. Machlin LJ, Pearson PB & Denton CA The utilization of sulfate sulfur for the synthesis of taurine in the developing chick embryo. J. Biol. Chem 212, 469–475 (1955). [PubMed: 13233249]
15. Chapeville F & Fromageot P Formation de sulfite, d'acide cystéique et de taurine à partir de sulfate par l'oeuf embryonné. Biochimica et biophysica acta 26, 538–558 (1957). [PubMed: 13499413]

16. Bennett NStudy of yolk-sac endoderm organogenesis in the chick using a specific enzyme (cysteine lyase) as a marker of cell differentiation. J. Embryol. Exp. Morphol29, 159–174 (1973). [PubMed: 4722150]
17. Fisher J-L, Smith J, Chapeville F & Dubois R La localisation de la cystéine lyase et le plan d'organisation du développement embryonnaire chez divers représentants de Vertébrés. Biol Cell 46, 291–300 (1982).
18. Tolosa EA, Chepurnova NK, Khomutov RM & Severin ES Reactions catalysed by cysteine lyase from the yolk sac of chicken embryo. Biochim. Biophys. Acta 171, 369–371 (1969). [PubMed: 5813025]
19. Ripps H & Shen W Review: taurine: a "very essential" amino acid. Mol. Vis 18, 2673–2686 (2012). [PubMed: 23170060]
20. Sturman JATaurine in development. J. Nutr118, 1169–1176 (1988). [PubMed: 3054019]
21. Cavallini D, Scandurra R, Dupre S, Santoro L & Barra D A new pathway of taurine biosynthesis. Physiol. Chem. Phys 8, 157–160 (1976). [PubMed: 981355]
22. Veeravalli Set al.Flavin-Containing Monooxygenase 1 (FMO1) Catalyzes the Production of Taurine from Hypotaurine. Drug Metab. Dispos (2020). doi:10.1124/dmd.119.089995
23. Norberg S, Powell TL & Jansson T Intrauterine growth restriction is associated with a reduced activity of placental taurine transporters. Pediatr. Res 44, 233–238 (1998). [PubMed: 9702920]
24. Heller-Stilb Bet al.Disruption of the taurine transporter gene (taut) leads to retinal degeneration in mice. FASEB J16, 231–233 (2002). [PubMed: 11772953]
25. Green ML & Karp PD A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. BMC Bioinformatics 5, 76 (2004). [PubMed: 15189570]
26. Nagy LG, Merényi Z, Hegedüs B & Bálint B Novel phylogenetic methods are needed for understanding gene function in the era of mega-scale genome sequencing. Nucleic acids research (2020). doi:10.1093/nar/gkz1241
27. Percudani R & Peracchi A The B6 database: a tool for the description and classification of vitamin B6-dependent enzymatic activities and of the corresponding protein families. BMC Bioinformatics 10, 273 (2009). [PubMed: 19723314]
28. Antin PB, Yatskievych TA, Davey S & Darnell DK GEISHA: an evolving gene expression resource for the chicken embryo. Nucleic Acids Res 42, D933–7 (2014). [PubMed: 24150938]
29. Majtan Tet al.Domain organization, catalysis and regulation of eukaryotic cystathionine beta-synthases. PLoS ONE9, e105290 (2014). [PubMed: 25122507]
30. Meier M, Janosik M, Kery V, Kraus JP & Burkhard P Structure of human cystathionine beta-synthase: a unique pyridoxal 5'-phosphate-dependent heme protein. EMBO J 20, 3910–3916 (2001). [PubMed: 11483494]
31. Taoka S, West M & Banerjee R Characterization of the heme and pyridoxal phosphate cofactors of human cystathionine beta-synthase reveals nonequivalent active sites. Biochemistry 38, 2738–2744 (1999). [PubMed: 10052944]
32. Paul BD & Snyder SH H_2S signalling through protein sulfhydration and beyond. Nat. Rev. Mol. Cell Biol 13, 499–507 (2012). [PubMed: 22781905]
33. Agnello G, Chang LL, Lamb CM, Georgiou G & Stone EM Discovery of a substrate selectivity motif in amino acid decarboxylases unveils a taurine biosynthesis pathway in prokaryotes. ACS Chem. Biol 8, 2264–2271 (2013). [PubMed: 23972067]
34. Burt DWet al.The Chicken Gene Nomenclature Committee report. BMC Genomics10Suppl 2, S5 (2009).
35. Hamburger V & Hamilton HL A series of normal stages in the development of the chick embryo. 1951. Dev. Dyn 195, 231–272 (1992). [PubMed: 1304821]
36. Sentenac A & Fromageot P La sérinehydrolyase de l'oiseau mise en évidence dans l'embryon et mécanisme d'action. Biochimica et Biophysica Acta (BBA) - Specialized Section on Enzymological Subjects 81, 289–300 (1964).
37. Bergsten JA review of long-branch attraction. Cladistics21, 163–193 (2005).
38. Haslewood GABile salts of germ-free domestic fowl and pigs. Biochem. J123, 15–18 (1971). [PubMed: 5128663]

39. Tomaselli Set al.NMR-based modeling and binding studies of a ternary complex between chicken liver bile acid binding protein and bile acids. Proteins69, 177–191 (2007). [PubMed: 17607743]
40. Haslewood GABile salt evolution. J. Lipid Res8, 535–550 (1967). [PubMed: 4862128]
41. Czuba B & Vessey DA Purification and characterization of cholyl-CoA: taurine N-acetyltransferase from the liver of domestic fowl (Gallus gallus). Biochem. J 195, 263–266 (1981). [PubMed: 7306052]
42. Noble RC & Cocchi M Lipid metabolism and the neonatal chicken. Prog. Lipid Res 29, 107–140 (1990). [PubMed: 2093896]
43. Yadgary L, Kedar O, Adepeju O & Uni Z Changes in yolk sac membrane absorptive area and fat digestion during chick embryonic development. Poult. Sci 92, 1634–1640 (2013). [PubMed: 23687160]
44. Thurston JH, Hauhart RE & Naccarato EF Taurine: possible role in osmotic regulation of mammalian heart. Science 214, 1373–1374 (1981). [PubMed: 7313699]
45. Wu J-Yet al.Mechanism of neuroprotective function of taurine. Adv. Exp. Med. Biol643, 169–179 (2009). [PubMed: 19239147]
46. Ackerman RAPhysiological and ecological aspects of gas exchange by sea turtle eggs. Am. Zool20, 575–583 (1980).
47. Metcalfe J, McCutcheon IE, Francisco DL, Metzenberg AB & Welch JE Oxygen availability and growth of the chick embryo. Respir. Physiol 46, 81–88 (1981). [PubMed: 7335987]
48. Germs ACHydrogen sulphide production in eggs and egg products as a result of heating. J. Sci. Food Agric24, 7–16 (1973). [PubMed: 4735258]
49. Gunnison AFSulphite toxicity: A critical review of in vitro and in vivo data. Food Cosmet. Toxicol19, 667–682 (1981). [PubMed: 6171492]
50. Schnell S & Mendoza C Closed Form Solution for Time-dependent Enzyme Kinetics. J. Theor. Biol 187, 207–212 (1997).
51. Ashkenazy Het al.FastML: a web server for probabilistic reconstruction of ancestral sequences. Nucleic Acids Res40, W580–4 (2012). [PubMed: 22661579]
52. Antin PB, Pier M, Sesepasara T, Yatskievych TA & Darnell DK Embryonic expression of the chicken Krüppel-like (KLF) transcription factor gene family. Dev. Dyn 239, 1879–1887 (2010). [PubMed: 20503383]
53. Kery V, Poneleit L & Kraus JP Trypsin cleavage of human cystathionine beta-synthase into an evolutionarily conserved active core: structural and functional consequences. Arch. Biochem. Biophys 355, 222–232 (1998). [PubMed: 9675031]
54. Chiku Tet al.H2S biogenesis by human cystathionine gamma-lyase leads to the novel sulfur metabolites lanthionine and homolanthionine and is responsive to the grade of hyperhomocysteinemia. J. Biol. Chem284, 11601–11612 (2009). [PubMed: 19261609]
55. Salsi Eet al.Exploring O-acetylserine sulfhydrylase-B isoenzyme from Salmonella typhimurium by fluorescence spectroscopy. Arch. Biochem. Biophys505, 178–185 (2011). [PubMed: 20937239]

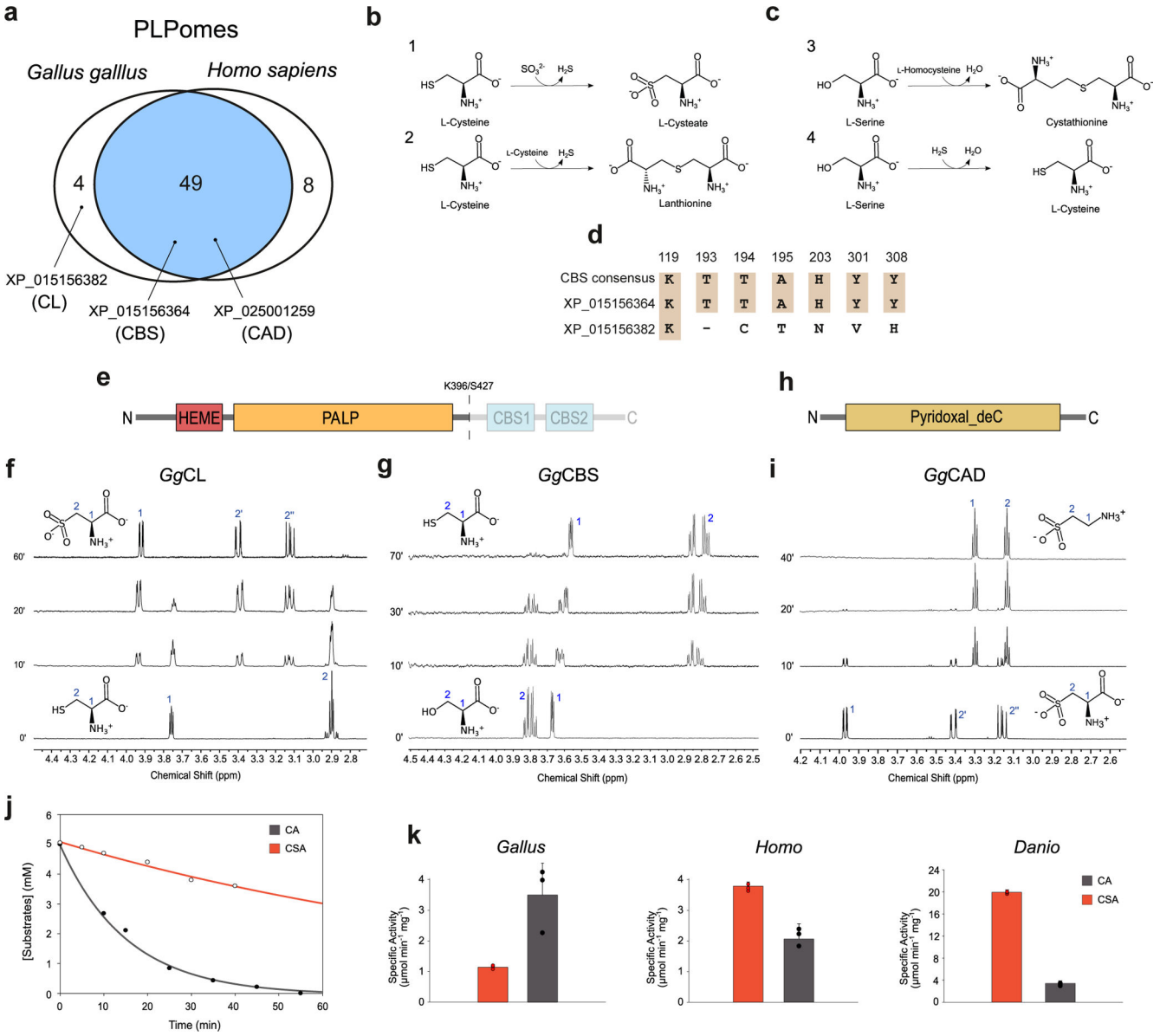


Fig. 1. Identification of the genes involved in sulfonated amino acids biosynthesis in *Gallus gallus*.

a, Venn diagram of the *in silico* comparison of *Gallus gallus* and *Homo sapiens* PLPomes summarizing the numbers of shared and unique genes. Accession numbers of the proteins identified as cysteine lyase (CL), cystathionine beta-synthase (CBS), cysteic acid decarboxylase (CAD) are indicated. **b**, Reactions catalyzed by CL: synthesis of cysteate (1) and synthesis of lanthionine (2). **c**, Reactions catalyzed by CBS: synthesis of cystathionine (3) and conversion of serine into cysteine via addition of hydrogen sulfide (4). **d**, Conservation of the catalytic lysine (K119) and non conservative substitutions of active site residues of *Gallus gallus* CBS-like (XP_015156382). Numeration is according to the human CBS sequence. **e,** *Gg*CL and *Gg*CBS domain composition; the dashed line indicates gene truncation for recombinant protein expression in *Gg*CL (K396) and *Gg*CBS (S427). **f**, Time-resolved ^{1}H NMR spectra of cysteine (5 mM) in the presence of Na_2SO_3 (7 mM) and

*Gg*CL (1 μM), showing complete conversion into cysteic acid. (**g**) Time-resolved ^{1}H NMR spectra of serine (5 mM) in the presence of Na_2S (30 mM) and *Gg*CBS (4 μM), showing complete conversion into cysteine. **h**, *Gg*CAD domain composition. (**i**) Time-resolved ^{1}H NMR spectra of cysteic acid (5 mM) in the presence of *Gg*CAD (1 μM), showing complete conversion into taurine. **j**, Example of *Gg*CAD kinetics in the presence of cysteic acid (CA) or cysteine sulfinic acid (CSA); the black curve is the fitting of the experimental points with the integrated Michaelis-Menten equation[50] with K_M = 6.95 ± 3.23 mM, k_{cat} = 10.54 ± 3.46 s^{-1}. **k**, Specific activities of *Gallus*, *Homo*, and *Danio* CSAD orthologs with CA or CSA substrates. Data are means ± SDV of three independent experiments.

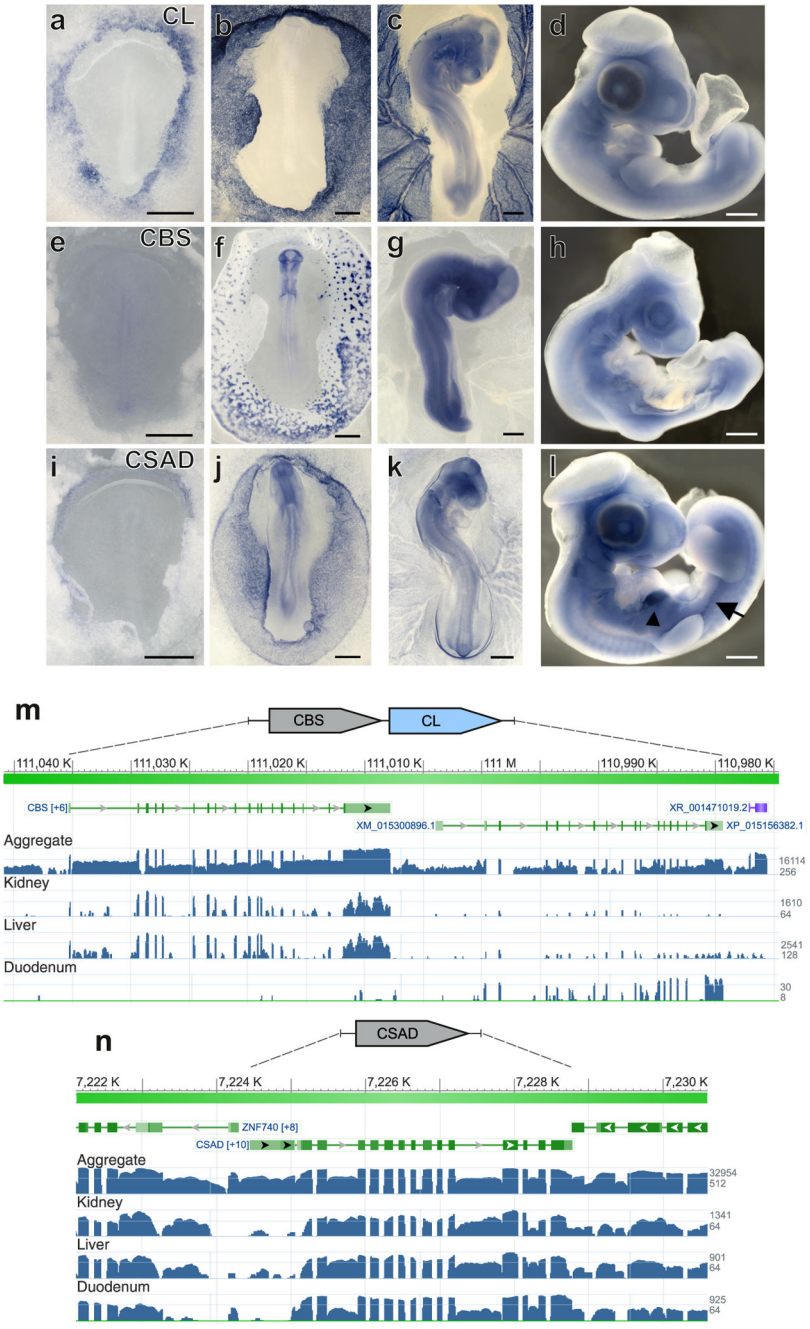


Fig. 2. CL, CBS, and CSAD genes are expressed since early stages of embryogenesis.

a-d, *In situ* hybridization analysis of CL expression in chick embryos at Hamburger-Hamilton developmental stages 4, 10, 18, and 24, sorted from left to right. **e-h**, *In situ* hybridization analysis of CBS expression at HH stages 4, 10, 18, and 24, sortexd from left to right. **i-l**, *In situ* hybridization analysis of CSAD expression at HH stages 4, 10, 18, and 24, sorted from left to right. Scale bars: 1mm (**a-c, e-g, i-k**) 5mm (**d, h, i**). **m**, NCBI Sequence-viewer representation of the genomic region on *Gallus gallus* chromosome 1 (annotation release 104) encompassing the CBS and CL genes. Gene exon structure is represented by green segments. Blue bars represent RNA-seq exon coverage (log2

scaled) for aggregate, kidney (SAMEA2201372), liver (SAMEA2201470), and duodenum (SAMN03376186) datasets. **n**, NCBI Sequence-viewer representation of the genomic region on *Gallus gallus* chromosome 33 encompassing the CSAD gene; tracks are as in panel m.

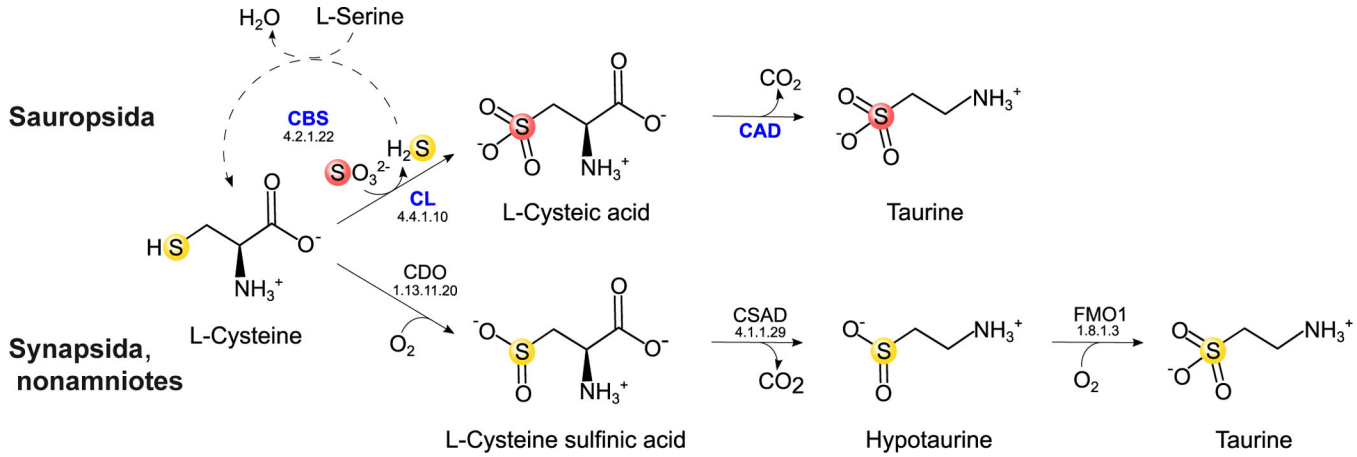


Fig. 3. The CL pathway for taurine biosynthesis.

The identified pathway for sulfonated amino acids biosynthesis is shown in comparison with the known pathway for taurine biosynthesis. The dashed line shows recycle of hydrogen sulfide into cysteine catalyzed by CBS. Cysteine and sulfite sulfur atoms are denoted in different colors to highlight the different sources of sulfur. Enzymes are indicated by EC numbers (if any) and protein abbreviations as follows: cystathionine beta-synthase (CBS), cysteine lyase (CL), cysteic acid decarboxylase (CAD), cysteine dioxygenase (CDO), cysteine sulfinic acid decarboxylase (CSAD), flavin-containing monooxygenase 1 (FMO1).

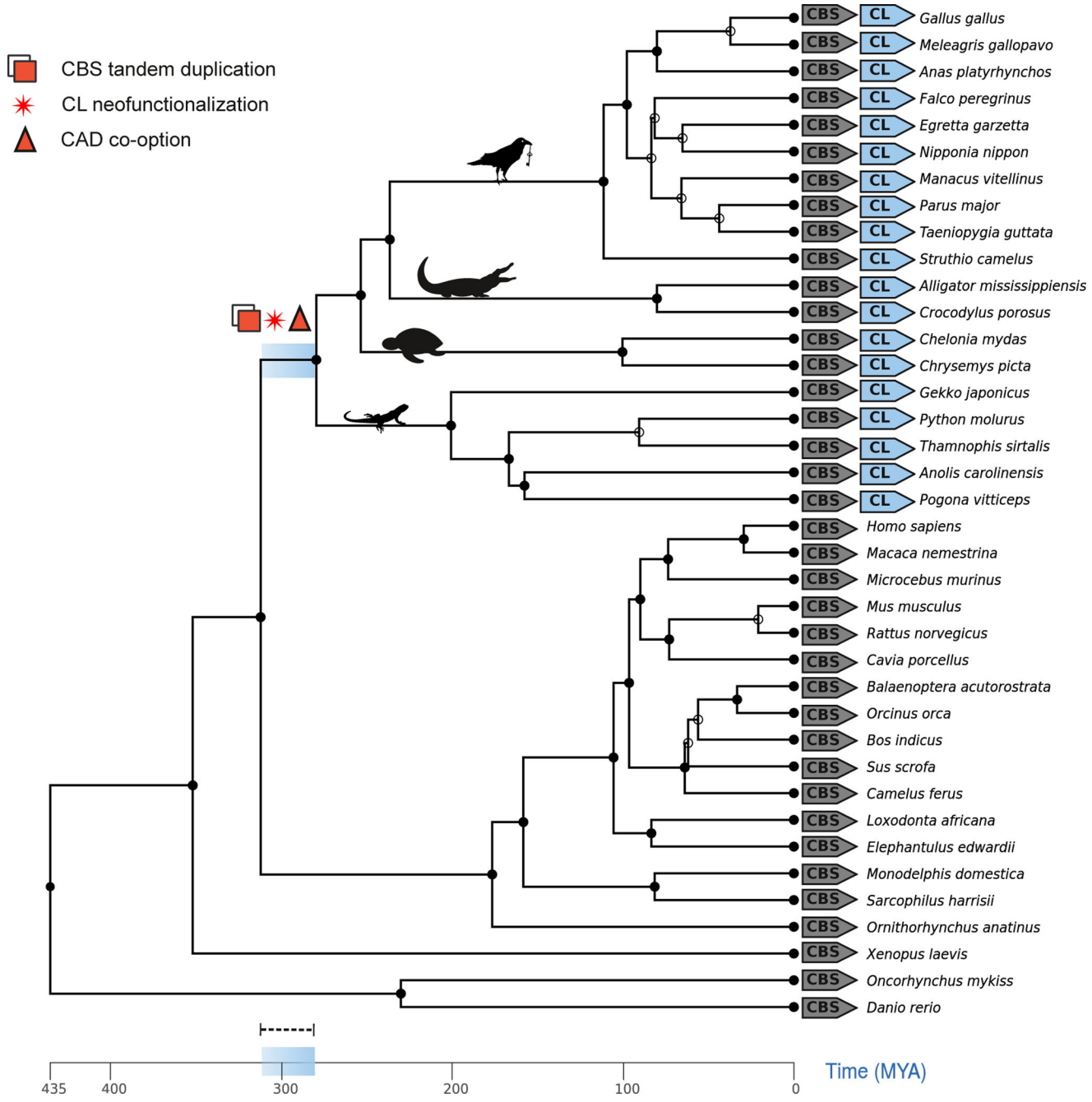


Fig. 4. Origin and conservation of the CL pathway in birds and reptiles.

Key evolutionary events in the origin of the metabolic pathway are mapped on a vertebrate chronogram. Phylogenetic relationships and divergence times are from TimeTree [1]; empty nodes correspond to unresolved relationships in NCBI taxonomy. The dashed line indicates uncertainty in dating the events along the branch and within the temporal boundaries delimited by blue segments. The configuration of the CBS locus in the different species is represented at the terminal nodes, showing conservation of CBS-CL synteny in sauropsids.